Size-Guided Multi-Seed Heuristic Method for Geometry Optimization of Clusters: Application to Benzene Clusters

Hiroshi Takeuchi*

Division of Chemistry, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

Corresponding author phone: +81-11-706-3533; Fax: +81-11-706-3501; e-mail: takehi@sci.hokudai.ac.jp

Abstract

Since searching for the global minimum on the potential energy surface of a cluster is very difficult, many geometry optimization methods have been proposed, in which initial geometries are randomly generated and subsequently improved with different algorithms. In this study, a size-guided multi-seed heuristic method is developed and applied to benzene clusters. It produces initial configurations of the cluster with *n* molecules from the lowest-energy configurations of the cluster with *n* - 1 molecules (seeds). The initial geometries are further optimized with the geometrical perturbations previously used for molecular clusters. These steps are repeated until the size *n* satisfies a predefined one. The method locates putative global minima of benzene clusters with up to 65 molecules. The performance of the method is discussed using the computational cost, rates to locate the global minima, and energies of initial geometries.

Keywords: global optimization · geometrical perturbation · initial geometry · growth sequence

Introduction

Nanoclusters have specific physical and chemical properties compared with the corresponding bulk material. To understand properties of a cluster, it is important to know the structure. It has been theoretically investigated with geometry optimization methods. Geometry optimization of the cluster is usually carried out using nondeterministic algorithms since there are huge number of local minima on the potential energy surface (PES).[1] When atomic clusters are compared with molecular ones, the latter shows more complicated PES than the former because of combinations of positions and orientations of molecules. The complication restricts the size of the clusters whose global-minimum geometries can be obtained to a few ten molecules. The purpose of this study is to improve this by developing a novel optimization method.

Optimization methods are divided into two groups, biased and unbiased algorithms.[2,3] Unbiased algorithms search for the global minimum from randomly generated geometries. Since these geometries are usually amorphous and significantly different from the corresponding optimal geometry, the potential energies are much higher than the lowest energy and must be minimized by a lot of geometrical perturbations and local optimizations. As improvements of the optimization algorithms, the following points are considered: (1) generation of more reasonable initial geometries than randomly generated ones; (2) development of efficient geometrical perturbations; and (3) reduction of the number of local minima by smoothing the PES. Devises on mutation and crossover operators in evolutionary algorithms are included in the point 2.[1,4] The present author proposed interior and surface operators as geometrical perturbations in the heuristic optimization method.[5] The point 3 is also considered in the previous studies[6,7] since it is easy to search for the global minimum on the smoothed PES. In the study by Pillardy et al.,[6] additional calculations are necessary to obtain the global minimum on the original PES from that on the smoothed PES. The basin hopping algorithm[7] smooths the PES with local optimizations, preserving the position of the global minimum. Other methods also use local optimizations to enhance the efficiency.[8-13] In

the study where the basin hopping algorithm is applied to Lennard-Jones atomic clusters,[7] 5 runs are performed from randomly generated geometries and 2 runs are executed with a seeding strategy. In the latter runs, the most stable configuration of the cluster with n atoms is used as a seed to produce the clusters with ($n \pm 1$) atoms; these initial geometries are generated from the seed by adding and subtracting an atom. Hence this strategy is related to the point 1. Other strategies reported for the point 1 use empirical information on the geometries.[14-16]

Since a geometry generated from a single seed corresponds to a point on the PES, exploration from it is restricted. Hence the single-seed method belongs to the category of biased algorithms. This may lead to the result that the global minimum is not found from the generated geometry. In practice, several unsuccessful cases are confirmed by Shao et al.[14] for Lennard-Jones atomic clusters. In this study, a new algorithm is proposed by improving the single-seed strategy. In the new algorithm, the lowest energy configurations of $(mol)_{n-1}$ are selected as seeds and several initial geometries of the n-molecule cluster $(mol)_n$ are generated from them to explore various spaces on the PES. The generated geometries are improved with the optimization method previously developed by the present author, the heuristic method combined with geometrical perturbations (HMGP).[17] These steps are repeated until the cluster size satisfies a predefined integer. This method called hereafter size-guided multi-seed heuristic method combined with geometrical perturbations (SGMS-HMGP) is applied to benzene clusters $(C_6H_6)_n$ ($n \leq 65$) to elucidate the performance. Many seeds would introduce unbiased property into the method. Hence the multi-seed strategy is expected to locate the global minima when the number of seeds is large.

Benzene Clusters and Potential Function

In this study, benzene clusters were selected as a test case since these were often investigated as described below. Optimal geometries of the clusters are reported[17-24] employing 3 intermolecular potentials developed by Williams and Starr (WS model),[25] Jorgensen and Severance (OPLS-AA

model),[26] and Bartolomei, Pirani, and Marques (BPM model).[20] The maximum number of molecules in the cluster is 30[17-19] for the OPLS-AA and WS models, and 25[20] for the BPM model. Other potentials are also used for only $(C_6H_6)_{13}$.[27] Accordingly the geometries of the clusters with more than 30 molecules have never been investigated. This indicates that it is difficult to locate the global minima of them.[17-20]

The BPM intermolecular potential is considered to be more reliable than the WS and OPLS-AA ones since the benzene dimer of the BPM model takes a T-shaped structure in accordance with the experiment whereas the WS and OPLS-AA models give more tilted structures.[20] Hence the BPM model is used in the present study. The potential energy of $(C_6H_6)_n$ is given in terms of the intermolecular interaction $V(i, j)$:

$$V(n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} V(i, j) \tag{1}$$

The interaction is expressed by the sum of electrostatic and non-electrostatic terms: $V(i, j) = V_e(i, j) + V_{non\text{-}e}(i, j)$. The former term is calculated with 12 negative charges (–0.04623 a.u.) on carbon atoms (two charges on each atom are separated by 1.905 Å) and 6 positive charges (0.09246 a.u.) on hydrogen atoms, and is presented by the distance r_{kl} between a charge k in molecule i and a charge l in molecule j:

$$V_e(i, j) = \sum_{k}^{18} \sum_{l}^{18} \frac{q_k q_l}{4\pi\varepsilon_0 r_{kl}} = \sum_{k}^{18} \sum_{l}^{18} \frac{c_{kl}}{r_{kl}} \tag{2}$$

where q and ε_0 represent electronic charges of sites and permittivity of vacuum, respectively. The non-electrostatic term is calculated with the distance r_{kl} between kth and lth atoms in molecules i and j as follows:

$$V_{\text{non-e}}(i,j)=\sum_{k}^{12}\sum_{l}^{12}\varepsilon_{kl}\left[\frac{m_{kl}}{n_{kl}(r_{kl})-m_{kl}}\left(\frac{r_{0,kl}}{r_{kl}}\right)^{n_{kl}(r_{kl})}-\frac{n_{kl}(r_{kl})}{n_{kl}(r_{kl})-m_{kl}}\left(\frac{r_{0,kl}}{r_{kl}}\right)^{m_{kl}}\right] - A_{\text{CH}}\left(\sum_{k\in\text{C}}\sum_{l\in\text{H}}\text{e}^{-\alpha_{\text{CH}}r_{kl}}+\sum_{k\in\text{H}}\sum_{l\in\text{C}}\text{e}^{-\alpha_{\text{CH}}r_{kl}}\right) \quad (3)$$

where ε_{kl} and $r_{0,kl}$, are the depth and equilibrium distance of the interaction, respectively. The last term is effective for C...H non-bonded interactions. The parameter n_{kl} is expressed as

$$n_{kl}(r_{kl})=\beta_{kl}+4.0\left(\frac{r_{kl}}{r_{0,kl}}\right)^{2} \quad (4)$$

The values of the potential parameters, c_{kl}, ε_{kl}, m_{kl}, $r_{0,kl}$, β_{kl}, A_{CH}, and α_{CH}, are listed in Table 1.

Geometry Optimization

Geometries of the small clusters ($n \le 5$) were easily optimized with a random search method. In the method, 200 geometries generated randomly were optimized with a limited memory quasi-Newton method.[28] The global minima of the n = 2, 3, 4, 5 clusters were located 200, 100, 30, and 30 times, respectively. The potential energies of them are listed in Table 2.

For the clusters with $n \ge 6$, SGMS-HMGP was used to obtain the global-minimum geometries. The flowchart of the method is shown in Figure 1. The following procedure is carried out: (1) The random search method is performed for $(C_6H_6)_5$ to yield seeds of $(C_6H_6)_6$ (part A in Figure 1). (2) Duplicate configurations are excluded from the seeds of $(C_6H_6)_n$ ($n \ge 6$) using the energies and rotational constants ($A \ge B \ge C$). Two geometries a and b are considered to be identical if the following conditions are satisfied: $|V_a(n) - V_b(n)|$/kJ mol^{-1} ≤ 0.01, $|A_a - A_b|/A_b \le 0.001$, $|B_a - B_b|/B_b \le 0.001$, and $|C_a - C_b|/C_b \le 0.001$. (3) The N_{seed} lowest-energy seeds of $(C_6H_6)_n$ are selected (part B in Figure 1). For each seed, a molecule is arbitrarily added on the surface and the geometry of the cluster is optimized with the quasi-Newton method.[28] (4) The resultant geometries of $(C_6H_6)_n$ are optimized with HMGP[17] (part C in Figure 1). (5) The size is increased by 1 and the second, third,

and fourth steps are performed until the size of the cluster satisfies a predefined value of 65.

In the step 3, seeds are selected with the energy-based criterion after the selection of parents in evolutionary algorithms.[4,29] However, this is not efficient for systems where similar seeds are often selected. In this case, the same optimized geometry is obtained many times. This may lower the ability of the method to locate the global minimum.

Using an evolutionary algorithm combined with local optimization, Pereira and Marques[30] investigated the relationship between the performance of the algorithm and the diversity of the population. The result shows that the diversity based on structural information is important to increase the efficiency of the algorithm. This suggests that the above case can be avoided through structural diversity of seeds. Combining the energy-based criterion with geometry-based criteria (local structures which are discussed later) would be useful.[31]

In the step 4, the interior (I), surface (S), and orientation (O) operators perturb geometries in this order.[17] Every geometry generated with the operator is optimized with the quasi-Newton method[28] as shown in Figure 1. The I operator moves m molecules with the highest potential energy to the surface of the sphere which takes the radius of $r_e/2$ (r_e denotes the equilibrium distance of the dimer 5.0 Å) and the center coincident with the center of mass of the cluster. Only outer molecules are selected as moved ones to reduce the number of the combinations calculated below. The energy of m molecules (the number of m is randomly selected from 1 to 5) is calculated with the following expression:

$$V_{m\,\mathrm{mol.}}(s_1, s_2, \cdots, s_m) = \sum_{j \neq s_1}^{n} V(s_1, j) + \sum_{j \neq s_1, s_2}^{n} V(s_2, j) + \cdots + \sum_{j \neq s_1, s_2, \cdots s_m}^{n} V(s_m, j) \qquad (5)$$

Here the numbering of molecules is expressed by s_1, s_2, …, s_m, From all the combinations of m molecules, the one with the highest potential energy is selected. When the energy of the cluster is lowered after the I operator followed by the local optimization, the geometry is updated. If the update does not occur during the last 10 operations, the S operator is carried out. This operator also

moves the highest-energy molecules. However, the positions of the moved molecules are different from those due to the I operator; the S operator selects the most stable positions on the surface of the cluster (for the detail of the stable positions, see ref [17]). The highest-energy, second highest-energy and third highest-energy molecules are separately moved in this order. Subsequently the number of the moved molecules increases at an interval of 1. When the energy-lowering is observed for the operation, the geometry is updated and the highest-energy molecule is moved again (the S operator returns to the initial step). If the S operator with 4 moved molecules does not improve the energy, the O operator is performed. That is, the orientations (Euler angles) of all molecules are randomly determined. When the O operator does not lower the energy during the last 10 operations, the calculation of the current geometry is terminated. The repetition times of the I, S, and O operators depend on the number of the updates of the geometries. The whole sum of the repetition times is shown later as computational cost. The number of the molecules moved with these operators were taken from the previous study on the benzene clusters.[17] As discussed later, the ability of the O operator to improve geometries was low for large clusters. This was found after all the geometries of the clusters with $n \leq 55$ were optimized. Hence it was not adopted for the clusters with $n \geq 56$.

In this work, 7 runs with $N_{seed} = 50$ and 4 runs with $N_{seed} = 100$ were performed; a run means the whole calculation from $(C_6H_6)_6$ to $(C_6H_6)_{65}$. The number of seeds was empirically determined by performing some test calculations. Since the numbers of the independent configurations of $(C_6H_6)_5$ and $(C_6H_6)_6$ were smaller than N_{seed}, those of initial geometries of $(C_6H_6)_6$ and $(C_6H_6)_7$ were also smaller than N_{seed}. However, those of the remaining clusters were equal to N_{seed}. Table 2 lists the lowest energies of the clusters with $6 \leq n \leq 65$ obtained with SGMS-HMGP. The corresponding geometries are deposited in the supporting information.

Discussion

Comparison with the Previous Study on BPM clusters

Bartolomei et al.[20] report the potential energies of the global minima and a few low-lying minima of $(C_6H_6)_n$ with n = 3, 13, 19. These values were equal to the corresponding ones obtained in the present study within 0.06 kJ mol^{-1}. The largest discrepancy was observed for the potential energy of the global minimum of $(C_6H_6)_{19}$; the values in the literature[20] and this study are –613.682 kJ mol^{-1} and –613.627 kJ mol^{-1}, respectively. This would be ascribed to the differences between the potential parameter values used in the previous[20] and present studies. The energy differences $\Delta V(n)$ between the global and local minima for $(C_6H_6)_n$ with n = 8, 9, 10 are also reported in the literature.[20] These were reproduced by the present method within 0.001 kJ mol^{-1}. However, the assignments of the local minima of $(C_6H_6)_n$ with n = 8, 10 in the literature are different from those in this study. In the previous study,[20] the minima with $\Delta V(8)$ = 2.188 kJ mol^{-1} and $\Delta V(10)$ = 1.535 kJ mol^{-1} are assigned to the second and third lowest-energy configurations, respectively whereas the present results show that these are assigned to the third and sixth lowest-energy configurations, respectively. Several local minima would be missing in these clusters.[20]

Performance of SGMS-HMGP

In the previous study[17] on the benzene clusters ($n \leq 30$) expressed by the WS model, a lot of randomly generated geometries were improved with HMGP to locate the global minima. These were later confirmed with the evolutionary algorithm.[19] Hence HMGP with randomly generated geometries is an excellent optimization method for benzene clusters with $n \leq 30$. In this study, for each of the n = 6 – 25, 30 BPM clusters, HMGP with 400 randomly generated geometries was performed to locate the global minimum at least twice whereas the clusters with n = 26 – 29 were omitted because of saving of computational time. The potential energies calculated with HMGP were equal to those in Table 2. Hence the global minima of the BPM clusters with $n \leq 30$ are considered to be reliable.

As mentioned in the introduction, it is expected that the multi-seed strategy introduces a certain unbiased property into the method. To clarify the property of the multi-seed strategy by comparing it with the single-seed one, 100 runs with $N_{seed} = 1$ were carried out for the clusters with $n \leq 40$. Each of these runs reproduced only the global minima for 3 to 13 cluster sizes. This indicates that SGMS-HMGP with 50 and 100 seeds has a certain unbiased nature.

For the n-molecule cluster, the performance of SGMS-HMGP and HMGP is examined using the average number of geometries required for searching for the global minimum $N_{per}(n)$. This number is equal to the number of local optimizations consuming most of computational time and thus represents computational effort to search for the global minimum. The number $N_{per}(n)$ is calculated using the equation:

$$N_{per}(n) = N_{all\ geom}(n)/N_{gm}(n) \quad (6)$$

Here $N_{all\ geom}(n)$ and $N_{gm}(n)$ denote the total number of geometries generated in the optimization of the cluster and the number of the initial geometries from which the global minimum is located, respectively. The results obtained for SGMS-HMGP and HMGP are shown in Figure 2 (the numerical data for SGMS-HMGP are deposited in the supporting information). For $n \geq 13$, the values of $N_{per}(n)$ of the former are smaller than those of the latter. Accordingly SGMS-HMGP is more efficient than HMGP for these clusters. Normally $N_{per}(n)$ increases with increasing size since the number of local minima on the PES exponentially increases with it. However, the value for SGMS-HMGP shows no significant increase for $30 \leq n \leq 65$. This indicates that SGMS-HMGP is useful for the large clusters which cannot be treated with HMGP.

The term $N_{all\ geom}(n)$ in eq. (6) is given by

$$N_{all\ geom}(n) = N_{seed}(n)N_{cost}(n) \quad (7)$$

where $N_{cost}(n)$ denotes the average number of geometries generated from an initial geometry and is used as computational cost per initial geometry in this study. Since a hit rate $R_{hit}(n)$ is expressed by $N_{gm}(n)/N_{seed}(n)$, eq. (6) is rewritten as

$$N_{per}(n) = N_{cost}(n)/R_{hit}(n) \quad (8)$$

Hence the performance is discussed with the computational cost and hit rate which are closely related to the energies of the initial geometries as described later. The cost and hit rate are shown in Figures 3 and 4, respectively (the data are deposited in the supporting information). In Figure 3, SGMS-HMGP gradually decreases the cost for $n \geq 12$ and significant decrease is found for $n = 56$ because the O operator is not carried out. However, $N_{cost}(n)$ of HMGP increases with increasing size and is higher than the corresponding one of SGMS-HMGP for $n \geq 12$. Figure 4 shows that most of the hit rates for SGMS-HMGP are higher than the corresponding rates for HMGP. Hence the performance of SGMS-HMGP is determined by the low computational cost and high hit rate.

The above results can be explained in terms of the initial geometries. Figure 5 shows the potential energies of them relative to the global-minimum ones; $\Delta V_{ini}(n) = V_{ini}(n) - V_{gm}(n)$ where $V_{ini}(n)$ and $V_{gm}(n)$ mean the potential energies of the initial and global-minimum geometries of $(C_6H_6)_n$, respectively. The values obtained with SGMS-HMGP are smaller than those with HMGP. This indicates that the initial geometries of SGMS-HMGP take more efficient packing than those of HMGP. Hence the number of the local optimizations required for the former is smaller than that for the latter. This is consistent with the discussion on the computational cost.

The hit rate is also related to the initial energies. The present method adopts the monotonic descent algorithm. Hence search spaces on the PES are restricted by the initial geometries since spaces with energies higher than the initial energies are prohibited. The differences shown in Figure 5 indicates that the search spaces of SGMS-HMGP are smaller than those of HMGP. This may increase the hit rate for SGMS-HMGP compared with that for HMGP, in agreement with the above discussion. Consequently the energy lowering of initial geometries enhances the performance of SGMS-HMGP through the low computational cost and high hit rate.

The relationship between performance of optimization methods and initial energies would hold for other methods with monotonic descent algorithms. Hence a key factor for the improvement of the algorithms is to generate initial geometries with low potential energies.

Two features are also detected in Figure 5; (1) there are striking peaks at the sizes of 13, 19, 34, 41, 43, and 57 in the result of SGMS-HMGP; and (2) the relative energy of SGMS-HMGP tends to slowly increases with increasing size. The feature 2 suggests that the size-guided multi-seed algorithm is more suitable for geometry optimization of large clusters than other methods with randomly generated geometries. The feature 1 is explained by the fact that these sizes correspond to the magic numbers clarified later.

The global minima of the clusters with $n > 30$ are obtained with SGMS-HMGP. To verify them and examine the efficiency of the method, application of other methods to these clusters would be necessary.

Geometrical Perturbations in SGMS-HMGP

To elucidate the performance of the geometrical perturbations, energy lowering due to each perturbation was examined. Figure 6 shows the values of $\Delta V_{gp}(n) = V_{after\ gp}(n) - V_{before\ gp}(n)$ where $V_{after\ gp}(n)$ and $V_{before\ gp}(n)$ mean the potential energies obtained after and before the geometrical perturbation followed by the local optimization, respectively. Three features are found for the results of SGMS-HMGP (Figure 6a): (1) for $n \leq 16$, the I operator lowers the potential energies more than the other operators; (2) for $n \geq 17$, the energy decrease due to the S operator is larger than that of the I operator; and (3) the O operator has little or no effect on the energy lowering for $n \geq 30$. Hence the O operator is not necessary for geometry optimization of large clusters. For HMGP (Figure 6b), the efficiency of the I operator is much higher than that of the S operator and the O operator contributes to location of the global minima as found for the WS benzene clusters.[17] These results show that the algorithm for generating initial geometries considerably affects the performance of the geometrical perturbations. A significant difference between the two methods is

found for the energy lowering due to the I operator since because packing of the initial geometries generated with the size-guided multi-seed algorithm is more efficient than that of randomly generated geometries.

Stepwise Increase of Size in Size-Guided Algorithm

The cluster size increases one by one in SGMS-HMGP. Even though one would like to treat only the *n*-molecule cluster, the present method requires beginning from the 6-molecule cluster. Hence the computational time of HMGP for $(C_6H_6)_n$ is compared with that of SGMS-HMGP required for locating the global minima of $(C_6H_6)_6$ to $(C_6H_6)_n$ in Figure 7; the time averaged over all the hits is used in the figure. The time of SGMS-HMGP is comparable to that of HMGP. Hence SGMS-HMGP and HMGP are useful for the single size calculation ($n \leq 30$) but the former is superior to the latter for larger sizes.

The increase in size Δn adopted in the algorithm is set to be 1 but any positive integer can be used for it. As the number of Δn gets larger and larger, SGMS-HMGP reaches a predefined cluster size quickly. However, energies of initial geometries generated with $\Delta n \geq 2$ would be larger than those with $\Delta n = 1$ since molecules added to the seed are arbitrarily placed on the cluster surface. Consequently, the performance decreases by increasing the number of Δn. At present, strategies satisfying the quickness and excellent performance are not found.

Growth Sequence of BPM Benzene Clusters

To understand structural features of the clusters, the molecule closest to the center of mass of the cluster was selected as an origin and distances between the centers of mass of the origin molecule and the other molecules were calculated.[20] The results are shown in Figure 8. The intermolecular distances less than 7 Å show formation of the first shell around the origin molecule. The second and third shells occur in the clusters with intermolecular distances larger than 7 and 12 Å, respectively. The borderlines of 7 and 12 Å are not distinct since the definition of the shells is unclear for distorted geometries observed for the benzene clusters as shown later.

The relative stability of the cluster is calculated with second energy difference:

$$\Delta_2 E_n = V_{gm}(n + 1) + V_{gm}(n - 1) - 2\, V_{gm}(n) \quad (9)$$

The values of $\Delta_2 E_n$ are shown in Figure 9. The results for $n \leq 25$ are in good agreement with those in the literature (Figure 7b in ref [20]). Since the cluster size is extended to 65, new magic numbers with $\Delta_2 E_n \geq 2$ kJ mol^{-1} are obtained; n = 26, 28, 30, 34, 41, 44, 48, 55, 57, 62. The magic numbers might be explained in terms of formation of stable local structures in the cluster. These are defined as follows; each of them is formed by a central molecule and 12 molecules surrounding it and the distances between the central and surrounding molecules are smaller than 7 Å. The number of the local structures $N_{lo}(n)$ in the global-minimum geometry is shown in Figure 10 together with the corresponding value for the Lennard-Jones cluster. The result for the benzene cluster is similar to that for the Lennard-Jones cluster. Hence the benzene clusters take the shell-by-shell[32] growth sequence observed for the Lennard-Jones ones. The magic numbers of 34, 55, and 57 are not coincident with the formation of the local structure since the increase in $N_{lo}(n)$ is not observed for these sizes. It was found that many local structures in the benzene clusters were deviated from those in the Lennard-Jones clusters. The stability of them depends on the deviations. Hence the number of $N_{lo}(n)$ may not be directly related to the relative stability of the clusters and thus the magic numbers. The deviations also affect the whole structures of the benzene clusters. Spherical structures are observed for the Lennard-Jones clusters with 13 and 55 atoms whereas the corresponding benzene clusters take prolate and oblate shapes, respectively since the asymmetry parameters $\kappa = (2B - A - C)/(A - C)$[18] calculated from the rotational constants (A, B, and C) are −1.00 and 0.62, respectively. The asymmetry parameter of the benzene cluster (Figure 11) shows a zig-zag line and this suggests that a lot of the global-minimum geometries of the benzene clusters irregularly grow with increasing size.

A transition from structures obeying shell-by-shell growth sequence to those in periodic solid states is an important property of clusters. This is not observed for the BPM benzene clusters since

no periodic feature emerges in them. However, acetylene clusters[33,34] and carbon dioxide clusters[35] show the transition at cluster sizes less than 40. The geometrical differences between benzene molecule and two linear molecules must be related to the transition but understanding of it in terms of molecular shapes is lacking.[36]

As described above, calculations of structures of clusters are important to investigate the growth sequence, building principle, magic numbers, and structural transitions. However, possible cluster sizes treated with the existing methods are so small that these cannot calculate the above properties in wide range of the sizes. Further development of optimization methods for molecular clusters is required to elucidate them.

Conclusions

Using the potential reported by Bartolomei et al.,[20] geometry optimizations of benzene clusters were carried out with SGMS-HMGP. The putative global minima of the clusters with up to 65 molecules are reported. The maximum size of the clusters is ca. 2 times as large as that in the previous studies (25 or 30). This indicates that SGMS-HMGP is efficient for geometry optimization of benzene clusters. The performance of the method is enhanced by the energy lowering of initial geometries due to the side-guided multi-seed algorithm. Using many seeds is essential to search for the global minima. The benzene clusters with $n \leq 65$ show shell-by-shell[32] growth sequence.

The size-guided multi-seed algorithm can be used for other optimization problems because of the following reasons: (1) the method requires no prior information on problems (geometrical features); (2) the implementation of the algorithm in other methods is easy; and (3) the algorithm is efficient for large cluster. The present study uses the heuristic method combined with the geometrical perturbations to improve geometries. However, other algorithms such as evolutionary algorithm[4,13,29] and basin-hopping algorithm[7] are also used as strategies of geometrical improvements. It would be very interesting to apply SGMS-HMGP to other systems,

multicomponent Lennard-Jones atomic clusters,[37-49] water clusters,[19, 50-66] and off-lattice protein models.[67-84] These applications are helpful to evaluate the size-guided multi-seed algorithm. The results of other optimization methods for the BPM benzene clusters are also required to elucidate the efficiency of the present method.

Additional Supporting Information may be found in the online version of this article.

References

1. Hartke, B. *WIREs Comput. Mol. Sci.* **1**, *2011*, 879-887.

2. Yang, X., Cai, W. and Shao, X. *J. Comput. Chem.* **28**, *2007*, 1427-1433.

3. Shao, X., Yang, X. and Cai, W. *J. Comput. Chem.* **29**, *2008*, 1772-1779.

4. Marques, J. M. C. and Pereira, F. B. *J. Mol. Liq.* **20**, *2015*, 51-63.

5. Takeuchi, H. *J. Chem. Inf. Model.* **46**, *2006*, 2066-2070.

6. Pillardy, J., Liwo, A. and Scheraga, H. A. *J. Phys. Chem. A* **103**, *1999*, 9370-9377.

7. Wales, D. J. and Doye, J. P. K. *J. Phys. Chem. A* **101**, *1997*, 5111-5116.

8. Deaven, D. M., Tit, N., Morris, J. R. and Ho, K. M. *Chem. Phys. Lett.* **256**, *1996*, 195-200.

9. Lv, J., Wang, Y., Zhu, L. and Ma, Y. *J. Chem. Phys.* **137**, *2012*, 084104.

10. Marques, J. M. C., Pereira, F. B. and Leitão, T. *J. Phys. Chem. A* **112**, *2008*, 6079-6089.

11. Shao, X., Cheng, L. and Cai, W. *J. Comput. Chem.* **25**, *2004*, 1693-1698.

12. Schönborn S. E, Goedecker, S., Roy, S. and Oganov, A. R. *J. Chem. Phys.* **130**, *2009*, 144108.

13. Johnston, R. L. *Dalton Trans. 2003*, 4193-4207.

14. Shao, X., Jiang, H. and Cai, W. *J. Chem. Inf. Model.* **44**, *2004*, 193-199.

15. Xiang, Y., Jiang, H., Cai, W. and Shao, X. *J. Phys. Chem. A* **108**, *2004*, 3586-3592.

16. Xiang, Y., Cheng, L., Cai, W. and Shao, X. *J. Phys. Chem. A* **108**, *2004*, 9516-9520.

17. Takeuchi, H. *J. Chem. Inf. Model.* **47**, *2007*, 104-109.

18. Takeuchi, H. *J. Phys. Chem. A* **116**, *2012*, 10172-10181.

19. Llanio-Trujillo, J. L., Marques, J. M. C. and Pereira, F. B. *J. Phys. Chem. A* **115**, *2011*, 2130-2138.

20. Bartolomei, M., Pirani, F. and Marques J. M. C. *J. Comput. Chem.* **36**, *2015*, 2291-2301.

21. Pullan, W. J. *J. Chem. Inf. Comput. Sci.* **37**, *1997*, 1189-1193.

22. White, R. P., Niesse, J. A. and Mayne, H. R. A. *J. Chem. Phys.* **108**, *1998*, 2208-2218.

23. Williams, D. E. *Acta Crystallogr. A* **36**, *1980*, 715-723.

24. van de Waal, B. W. *Chem. Phys. Lett.* **123**, *1986*, 69-72.

25. Williams, D. E. and Starr, T. L. *Comput. Chem.* **1**, *1977*, 173-177.

26. Jorgensen, W. L. and Severance, D. L. *J. Am. Chem. Soc.* **112**, *1990*, 4768-4774.

27. Easter, D. C. *J. Clust. Sci.* **15**, *2004*, 33-45.

28. Liu, D. C. and Nocedal, J. *Math. Prog*. **45**, *1989*, 503-528.

29. Marques, J. M. C., Pereira, F. B., Llanio-Trujillo, J. L., Abreu, P. E., Aquilar, A., Pirani, F. and Bartolomei, M. *Phil. Trans. R. Soc. A* **375**, *2017*, 20160198.

30. Pereira, F. B. and Marques, J. M. C. *Evol. Intel*. **2**, *2009*, 121-140.

31. Dittner, M. and Hartke, B. *Comput. Theoret. Chem.* **1107**, *2017*, 7-13.

32. Teo, B. K. *J. Clust. Sci.* **25**, *2014*, 5-28.

33. Takeuchi, H. *J. Comput. Chem.* **31**, *2010*, 1699-1706.

34. Shuler, K. and Dykstra, C. E. *J. Phys. Chem. A* **104**, *2000*, 11522-11530.

35. Takeuchi, H. *J. Phys. Chem. A* **112,** *2008*, 7492-7497.

36. Hartke, B. *Angew. Chem. Int. Ed.* **41**, *2002*, 1468-1487.

37. Takeuchi, H. *Comput. Theor. Chem.* **1050**, *2014*, 68-73.

38. Takeuchi, H. *Chem. Phys.* **457**, *2015*, 106-113.

39. Doye, J. P. K. and Meyer, L. *Phys. Rev. Lett.* **95**, *2005*, 063401.

40. Cassioli, A., Locatelli, M. and Schoen, F. *Optim. Methods Software* **24**, *2009*, 819-835.

41. Marques, J. M. C. and Pereira, F. B. *Chem. Phys. Lett.* **485**, *2010*, 211-216.

42. Kolossváry, I. and Bowers, K. J. *Phys. Rev. E* **82**, *2010*, 056711.

43. Sicher, M., Mohr, S. and Goedecker, S. *J. Chem. Phys.* **134**, *2011*, 044106.

44. Tao, Y., Ruchu, X. and Wenqi, H. *J. Chem. Inf. Model.* **51**, *2011*, 572-577.

45. Rondina, G. G. and Da Silva, J. L. F. *J. Chem. Inf. Model.* **53**, *2013*, 2282-2298.

46. Wu, X., Sun, Y., Li, C. and Yang, W. *J. Phys. Chem. A* **116**, *2012*, 8218-8225.

47. Wu, X., Huang, C., Sun, Y. and Wu, G. *Chem. Phys.* **415**, *2013*, 69-75.

48. Wu, X., Sun, Y., Gao, Y. and Wu, G. *J. Mol. Model.* **19**, *2013*, 3119-3125.

49. Dieterich, J. M. and Hartke, B. *J. Comput. Chem.* **32**, *2011*, 1377-1385.

50. Kazachenko, S. and Thakkar, A. J. *J. Chem. Phys.* **138**, *2013*, 194302.

51. Kazachenko, S. and Thakkar, A. J. *Mol. Phys.* **108**, *2010*, 2187-2193.

52. Kazachenko, S. and Thakkar, A. J. *Chem. Phys. Lett.* **476**, *2009*, 120-124.

53. Pillardy, J., Olszewski, K. A. and Piela, L. *J. Mol. Struct*., **270**, *1992*, 277-285.

54. Tsai. C. J. and Jordan, K. D. *J. Phys. Chem*., **97**, *1993*, 5208-5210.

55. Sremaniak, L. S., Perera, L. and Berkowitz, M. L. *J. Chem. Phys*., **105**, *1996*, 3715-3721.

56. Niesse, J. A. and Mayne, H. R. *J. Comput. Chem*., **18**, *1997*, 1233-1244.

57. Wales, D. J. and Hodges, M. P. *Chem. Phys. Lett*., **286**, *1998*, 65-72.

58. Qian, J., Stöckelmann, E. and Hentschke, R. *J. Mol. Model*., **5**, *1999*, 281-286.

59. Hartke, B. *Z. Phys. Chem.*, **214**, *2000*, 1251-1264.

60. Guimarães, F. F., Belchior, J. C., Johnston, R. L. and Roberts, C. *J. Chem. Phys*., **116**, *2002*, 8327-8333.

61. Kabrede, H. and Hentschke, R. *J. Phys. Chem. B*, **107**, *2003*, 3914-3920.

62. Hartke, B. *Phys. Chem. Chem. Phys*., **5**, *2003*, 275-284.

63. Kazimirskil, J. K. and Buch, V. *J. Phys. Chem. A*, **107**, *2003*, 9762-9775.

64. James, T., Wales, D. J. and Hernández-Rojas, J. *Chem. Phys. Lett*., **415**, *2005*, 302-307.

65. Kabrede, H. *Chem. Phys. Lett*., **430**, *2006*, 336-339.

66. Bandow, B. and Hartke, B. *J. Phys. Chem. A* **110**, *2006*, 5809-5822.

67. Takeuchi, H. *Mol. Inf.* **36**, *2017*, 1600096.

68. Irbäck, A., Peterson, C., Potthast, F. and Sommelius, O. *J. Chem. Phys.* **107**, *1997*, 273-282.

69. Hsu, H. -P., Mehra, V. and Grassberger, P. *Phys. Rev. E* **68**, *2003*, 037703.

70. Liang, F. *J. Chem. Phys.* **120**, *2004*, 6756-6763.

71. Kim, S. -Y., Lee, S. B. and Lee, J. *Phys. Rev. E* **72**, *2005*, 011916.

72. Bachmann, M., Arkin, H. and Janke, W. *Phys. Rev. E* **71**, *2005*, 031906.

73. Huang, W. and Liu, J. *Biopolymers* **82**, *2006*, 93-98.

74. Chen, M. and Huang, W. *J. Zhejiang Univ. SCIENCE B* **7**, *2006*, 7-12.

75. Kim, J., Straub, J. E. and Keyes, T. *Phys. Rev. E* **76**, *2007*, 011913.

76. Zhang, C. and Ma, J. *Phys. Rev. E* **76**, *2007*, 036708.

77. Schnabel, S., Bachmann, M. and Janke, W. *Phys. Rev. Lett.* **98**, *2007*, 048103.

78. Lee, J., Joo, K., Kim, S. -Y. and Lee, J. *J. Comput. Chem.* **29**, *2008*, 2479-2484.

79. Arkin, H. *Phys. Rev. E* **78**, *2008*, 041914.

80. Zhang, C. and Ma, J. *J. Chem. Phys.* **130**, *2009*, 194112.

81. Kim, J., Straub, J. E. and Keyes, T. *Phys. Rev. Lett.* **97**, *2006*, 050601.

82. Dash, T. and Sahu, P. K. *J. Comput. Chem.* **36**, *2015*, 1060-1068.

83. Kalegari, D. H. and Lopes, H. S. Proc. 2013 IEEE Symposium on Differential Evolution (SDE), IEEE, 2013, 143-150.

84. Parpinelli, R. S., BeníItez, C. M. V., Cordeiro, J. and Lopes, H. S. *J. Mult.-Valued Logic & Soft Computing*, **22**, *2014*, 267-286.

Table 1. Potential parameters of interatomic interactions[a]

sites i and j	c_{ij}/kJ mol^{-1}	ε_{ij}/kJ mol^{-1}	$r_{0,ij}$/Å	β_{ij}	m_{ij}	A_{ij}/kJ mol^{-1}	α_{ij}/Å^{-1}
C, C	2.9693471[b]	0.3222609	4.073	9.0	6		
C, H	-5.9386942[b]	0.1929706	3.505	6.5	6	1833.221	2.9
H, H	11.8773884	0.1553413	3.099	9.0	6		

[a] The parameter values are taken from ref [20].

[b] In the calculation of the electrostatic term, the position of the positive charge is used for the C atom and the atomic position is used for the H atom.

Table 2. The lowest energies of the benzene clusters $(C_6H_6)_n$ expressed by the BPM potential, $V(n)$ (kJ mol^{-1}). The values for n = 2 – 5 are obtained with a random search method and the other values are obtained with the size-guided multi-seed heuristic method.

n	$-V(n)$	n	$-V(n)$	n	$-V(n)$	n	$-V(n)$
2	12.5	18	569.3	34	1231.6	50	1927.4
3	37.1	19	613.7	35	1271.2	51	1970.4
4	65.6	20	652.0	36	1314.7	52	2013.5
5	93.0	21	689.5	37	1358.2	53	2057.5
6	125.9	22	730.7	38	1402.8	54	2102.9
7	156.0	23	773.0	39	1446.6	55	2153.1
8	190.2	24	812.8	40	1491.6	56	2198.3
9	223.3	25	855.6	41	1537.3	57	2244.3
10	258.7	26	898.2	42	1578.6	58	2286.7
11	294.0	27	938.8	43	1623.6	59	2329.2
12	332.7	28	982.3	44	1667.4	60	2376.5
13	380.2	29	1021.9	45	1707.5	61	2421.9
14	415.5	30	1063.9	46	1752.4	62	2467.9
15	451.5	31	1104.0	47	1795.7	63	2508.7
16	490.6	32	1146.7	48	1841.1	64	2554.6
17	527.2	33	1189.8	49	1882.9	65	2600.8

Figure captions

Figure 1. Flowchart of the size-guided multi-seed heuristic method combined with geometrical perturbations. The A, B, and C parts consist of the random search, size-guided multi-seed, and heuristic algorithms, respectively.

Figure 2. The performance $N_{per}(n)$ of the optimization methods which is represented by the average number of geometries required to obtain the global minimum: closed circles, the size-guided multi-seed heuristic method; open circles, the heuristic method with randomly generated geometries.

Figure 3. The computational cost $N_{cost}(n)$ of the optimization methods which is represented by the number of generated geometries per initial geometry: closed circles, the size-guided multi-seed heuristic method; open circles, the heuristic method with randomly generated geometries.

Figure 4. The hit rate $R_{hit}(n)$ of the size-guided multi-seed heuristic method (solid circles) and the heuristic method with randomly generated geometries (open circles).

Figure 5. The relative potential energies $\Delta V_{ini}(n)$ of initial geometries; $\Delta V_{ini}(n) = V_{ini}(n) - V_{gm}(n)$ where $V_{ini}(n)$ and $V_{gm}(n)$ mean the potential energies of the initial and global-minimum geometries of $(C_6H_6)_n$, respectively. The values of the size-guided multi-seed heuristic method and the heuristic method with randomly generated geometries are shown by closed and open circles, respectively.

Figure 6. The energy lowering due to the geometrical perturbations followed with local optimizations; A, the size-guided multi-seed heuristic method; B, the heuristic method with randomly generated geometries. The differences$\Delta V_{gp}(n)$ between the potential energies obtained before and after each perturbation are plotted; closed circles, interior operator; open circles, surface operator; open square, orientation operator.

Figure 7. Comparison of the average computational time of the size-guided multi-seed heuristic method to obtain all the global minima of $(C_6H_6)_6$ to $(C_6H_6)_n$ (closed circles) with that of the

heuristic method with randomly generated geometries to obtain the global minimum of $(C_6H_6)_n$ (open circles).

Figure 8. Distribution of distances between the origin and other molecules of benzene clusters.

Figure 9. The relative stability $\Delta_2 E_n$ of the benzene clusters.

Figure 10. The number of the local structures $N_{lo}(n)$ in the global-minimum geometry of the benzene cluster (closed circles) together with the corresponding one of the Lennard-Jones cluster (open circles).

Figure 11. Asymmetry parameter κ of the global-minimum geometry of the benzene cluster.

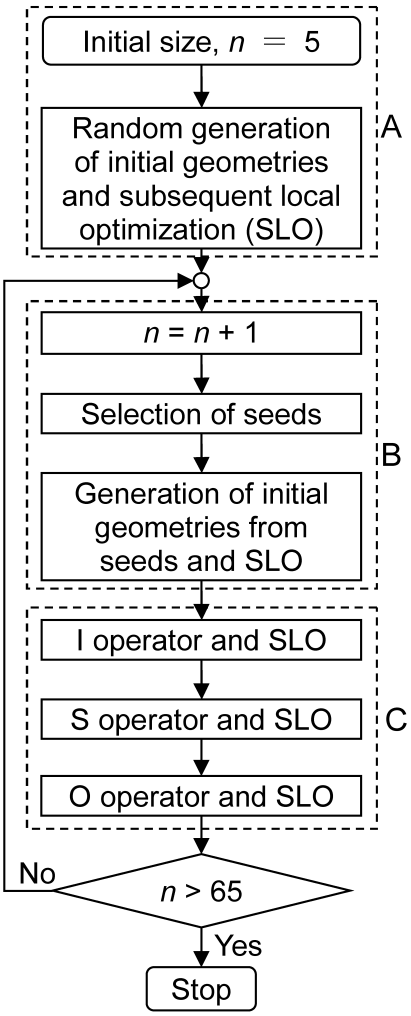
Initial size, n = 5
Random generation of initial geometries and subsequent local optimization (SLO)
A
n = n + 1
Selection of seeds
Generation of initial geometries from seeds and SLO
B
I operator and SLO
S operator and SLO
O operator and SLO
C
No
n > 65
Yes
Stop

Fig. 1

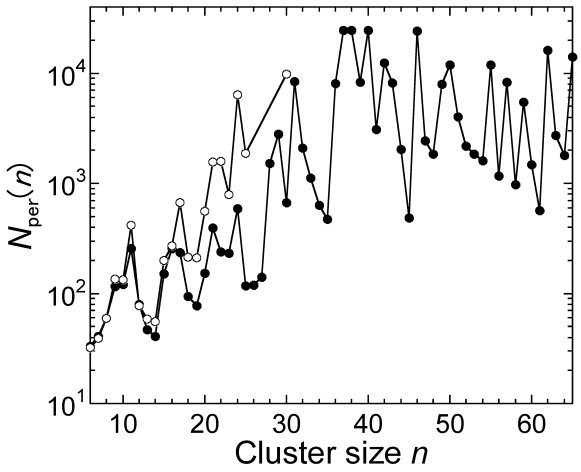
10^4
10^3
10^2
10^1
$N_{per}(n)$
10
20
30
40
50
60
Cluster size n

Fig. 2

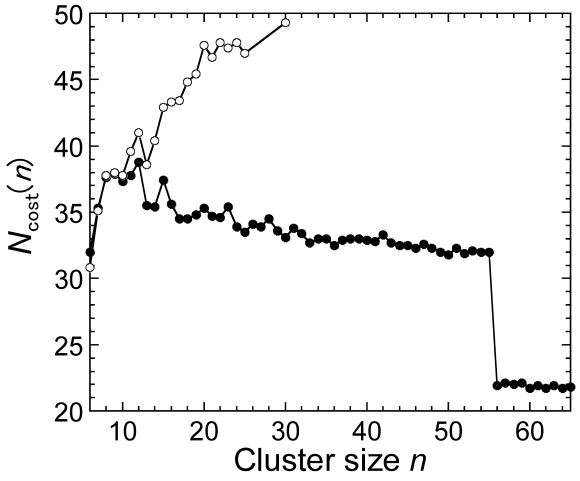
50
45
40
35
30
25
20
$N_{cost}(n)$
10
20
30
40
50
60
Cluster size n

Fig. 3

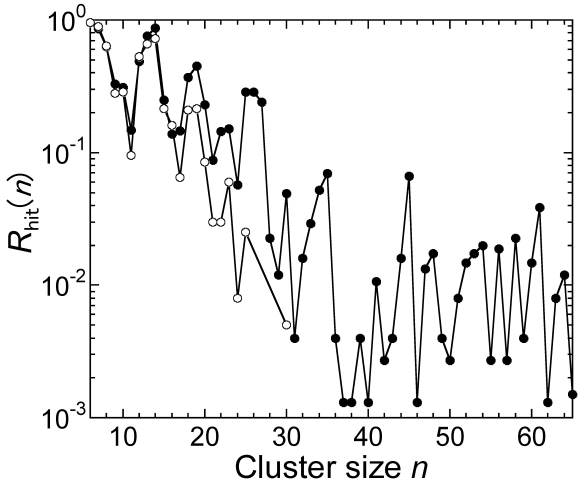
10^0
10^{-1}
10^{-2}
10^{-3}
$R_{hit}(n)$
10
20
30
40
50
60
Cluster size n

Fig. 4

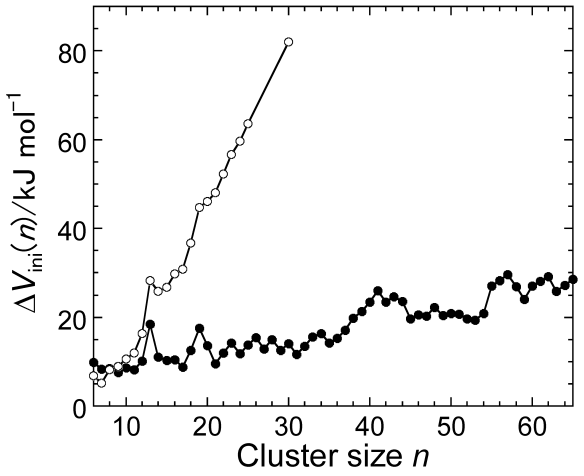
80
60
40
20
0
$\Delta V_{ini}(n)$/kJ mol^{-1}
10
20
30
40
50
60
Cluster size n

Fig. 5

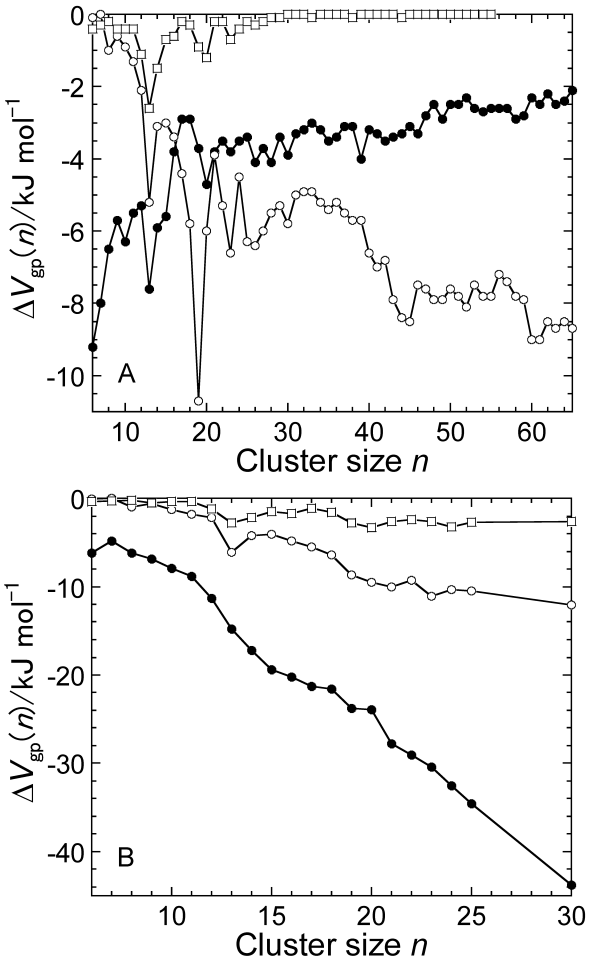
0
-2
-4
-6
-8
-10
$\Delta V_{gp}(n)$/kJ mol^{-1}
A
10
20
30
40
50
60
Cluster size n
0
-10
-20
-30
-40
$\Delta V_{gp}(n)$/kJ mol^{-1}
B
10
15
20
25
30
Cluster size n

Fig. 6

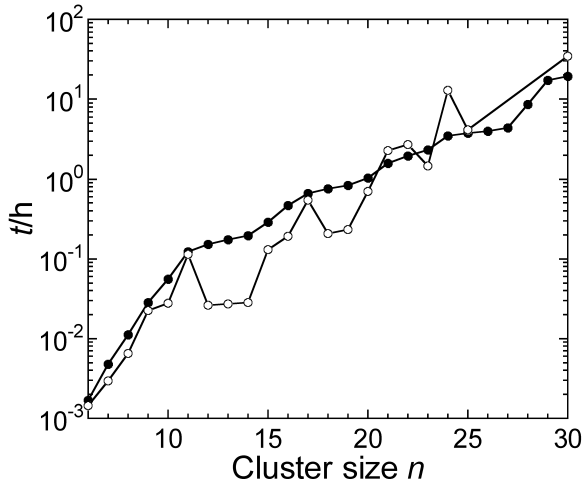
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
t/h
10
15
20
25
30
Cluster size n

Fig. 7

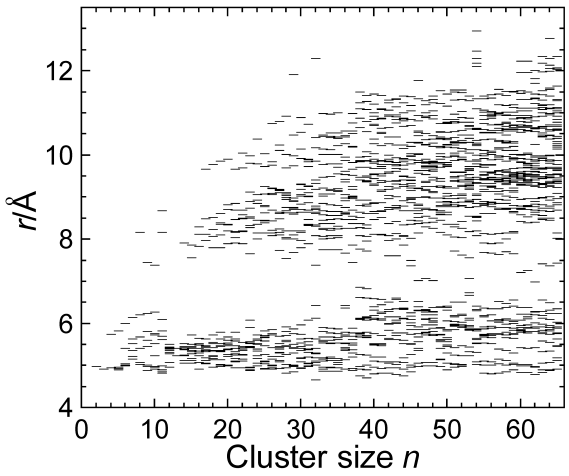
r/Å
12
10
8
6
4
0
10
20
30
40
50
60
Cluster size n

Fig. 8

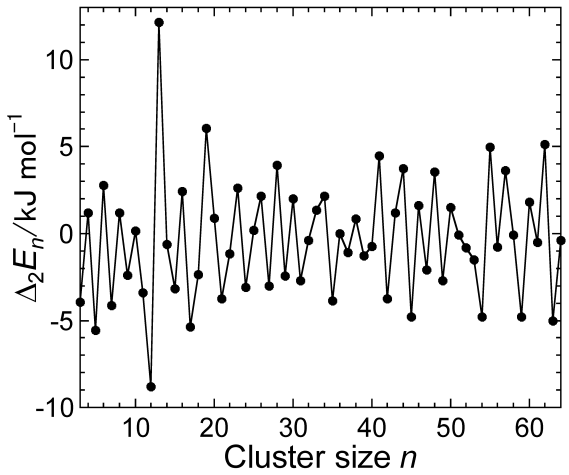
Δ2En/kJ mol-1
10
5
0
-5
-10
10
20
30
40
50
60
Cluster size n

Fig. 9

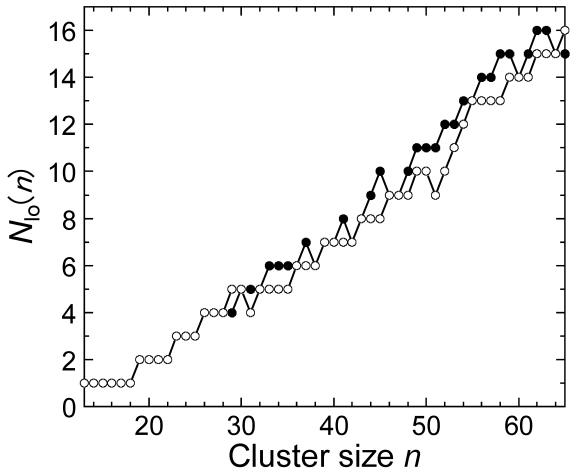

Fig. 10

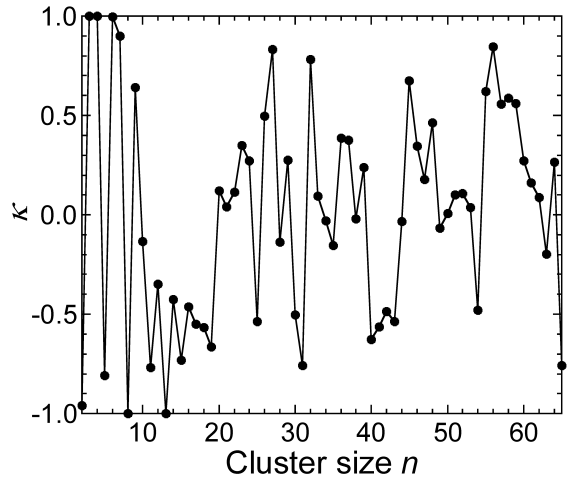

Fig. 11